

## 引用意見とその発言者を利用した意見情報の提示

佐藤 沙紀

論説型レポートの執筆や、政策判断のための世論調査を行うために、意見情報を有効に活用することが望まれている。本研究では、レポートテーマを表す情報要求に対し、複数の立場からの意見が Web 上に多数存在する場合、発言者の立場に着目し意見情報を抽出し、提示する手法を提案して、その有用性について実験を行った。ここでは、ニュース記事に含まれる引用部分の中に、ユーザの情報要求に適合する意見情報があると考え、その意見情報を引用意見と呼ぶことにする。また、引用意見の周辺には、その意見の発言者が存在すると考えられる。その発言者である、個人名や、組織名を立場とし、抽出した引用意見をこの立場に基づいて分類・提示することを試みる。まず、Web 上のニュースサイトを起点とし、NICT Web クローラを使用してニュース記事を収集する。次に、収集した記事中の引用部分“[…]”に着目し、その中に含まれる引用意見を抽出する。そして、引用意見の周囲に現れる発言者の個人名や組織名を、立場として判別することで、立場に基づいた意見情報の分類を試みる。

実験では、Web 上の検索結果をもとに作成したテストデータを用いて、引用意見の周辺に出現する個人名、組織名といった立場を抽出できるかを検証した。この時、立場となる語は固有表現であると考え、Cabocha の固有表現抽出機能を使用し、その中でも人物を表す <PERSON>、組織を表す <ORGANIZATION> を利用して、立場を判別することとした。立場の判別について評価したところ、『係助詞“は”から“[”までの距離が 15 字以内の場合、係助詞“は”の直前の語を抽出し、人名と組織名が含まれていた場合、人名を優先して抽出する』という手法が他の手法より有効であることがわかった。

次に、Web からクローラしたデータを用いて、抽出した引用意見の適合性について判断するために、実験を行った。また、適合した引用意見に対し、適切な立場が判別できているかの評価も行った。まず、事前にいくつかレポート作成用のテーマを準備し、そのテーマごとに実験に使用するクエリを設定し、クエリの拡張を行い関連語を得た。次に、設定したクエリ、またはその関連語を含む引用意見の抽出を行い、立場の判別を試みた。その結果、人の生活に関連が深く、意見交換が盛んに行われている話題に対して適合する引用意見を得られるということが明らかとなった。また、立場の判別が行えない場合として、複数の引用意見が連続して現れる事例を確認した。

今後の予定としては、さらなる精度向上のために広告を取り除くことを考えている。広告に含まれる引用部分と、引用意見を含む引用部分との差を明らかにすることで、広告を除くことができると考えている。また、汎用性の高い関連語を拡張すると適合しない意見が多く取れてしまうことから、関連語の絞り込み方法の改善についても検討していきたい。

(指導教員 関 洋平)